240D Cameron Fall 2004
Solutions to Final Exam

1.(a) Here

$$\ln L(\boldsymbol{\beta},\alpha) = \sum_i \ln f(y_i) = \sum_i \left\{ (\alpha-1)\ln y_i - \frac{y_i}{\exp(\mathbf{x}_i'\boldsymbol{\beta})} - \alpha \mathbf{x}_i'\boldsymbol{\beta} - \ln\Gamma(\alpha) \right\}$$

(b) Differentiation yields

$$\begin{aligned} \frac{\partial \ln L}{\partial \boldsymbol{\beta}} &= \sum_i \left(\frac{y_i}{\exp(\mathbf{x}_i'\boldsymbol{\beta})}\mathbf{x}_i - \alpha\mathbf{x}_i \right) = \sum_i \left(\frac{y_i - \alpha\exp(\mathbf{x}_i'\boldsymbol{\beta})}{\exp(\mathbf{x}_i'\boldsymbol{\beta})}\mathbf{x}_i \right) = \mathbf{0} \\ \frac{\partial \ln L}{\partial \alpha} &= \sum_i \left(\ln y_i - \mathbf{x}_i'\boldsymbol{\beta} - \frac{\Gamma'(\alpha)}{\Gamma(\alpha)} \right) = 0. \end{aligned}$$

(c) Easiest to derive the outer product of the gradient estimate $\widehat{\mathbf{B}}^{-1}$. This yields for $\boldsymbol{\theta} = [\boldsymbol{\beta}' \ \alpha]'$.

$$\widehat{\mathbf{V}}[\widehat{\boldsymbol{\theta}}] = \begin{bmatrix} \sum_i \left(\frac{y_i - \widehat{\alpha}\exp(\mathbf{x}_i'\widehat{\boldsymbol{\beta}})}{\exp(\mathbf{x}_i'\widehat{\boldsymbol{\beta}})}\right)^2 \mathbf{x}_i\mathbf{x}_i' & \sum_i \left(\ln y_i - \mathbf{x}_i'\widehat{\boldsymbol{\beta}} - \frac{\Gamma'(\widehat{\alpha})}{\Gamma(\widehat{\alpha})}\right)\left(\frac{y_i - \widehat{\alpha}\exp(\mathbf{x}_i'\widehat{\boldsymbol{\beta}})}{\exp(\mathbf{x}_i'\widehat{\boldsymbol{\beta}})}\right)\mathbf{x}_i \\ \sum_i \left(\frac{y_i - \widehat{\alpha}\exp(\mathbf{x}_i'\widehat{\boldsymbol{\beta}})}{\exp(\mathbf{x}_i'\widehat{\boldsymbol{\beta}})}\right)\left(\ln y_i - \mathbf{x}_i'\widehat{\boldsymbol{\beta}} - \frac{\Gamma'(\widehat{\alpha})}{\Gamma(\widehat{\alpha})}\right)\mathbf{x}_i & \sum_i \left(\ln y_i - \mathbf{x}_i'\widehat{\boldsymbol{\beta}} - \frac{\Gamma'(\widehat{\alpha})}{\Gamma(\widehat{\alpha})}\right)^2 \end{bmatrix}^{-1}$$

Or can use Hessian which $-\widehat{\mathbf{A}}^{-1}$ yields after some algebra yields

$$\widehat{\mathbf{V}}[\widehat{\boldsymbol{\theta}}] = \begin{bmatrix} \sum_i \left(\frac{y_i}{\exp(\mathbf{x}_i'\widehat{\boldsymbol{\beta}})}\right)\mathbf{x}_i\mathbf{x}_i' & \sum_i \mathbf{x}_i \\ \sum_i \mathbf{x}_i & \sum_i \left(\frac{\Gamma'(\widehat{\alpha})}{\Gamma(\widehat{\alpha})} - \frac{\Gamma'(\widehat{\alpha})^2}{\Gamma(\widehat{\alpha})^2}\right) \end{bmatrix}^{-1}$$

(d) This is a tough one. In general the MLE will be inconsistent.
Here there is some hope that this may not be the case, since $\mathrm{E}[\partial \ln L/\partial\boldsymbol{\beta}] = \mathbf{0}$ requires only correct specification of the mean (then $\mathrm{E}\left[\sum_i \frac{y_i - \alpha\exp(\mathbf{x}_i'\boldsymbol{\beta})}{\exp(\mathbf{x}_i'\boldsymbol{\beta})}\mathbf{x}_i\right] = 0$).
But $\mathrm{E}[\partial \ln L/\partial\alpha] = 0$ requires much stronger assumption that $\mathrm{E}[\ln y_i] = \mathbf{x}_i'\boldsymbol{\beta} + \frac{\Gamma'(\alpha)}{\Gamma(\alpha)}$ (then $\mathrm{E}\left[\sum_i \left(\ln y_i - \mathbf{x}_i'\boldsymbol{\beta} - \frac{\Gamma'(\alpha)}{\Gamma(\alpha)}\right)\right] = 0$).
This fails and the two equations jointly estimated will yield inconsistent estiamtes.
So both $\widehat{\boldsymbol{\beta}}$ and $\widehat{\alpha}$ will be inconsistent.

(e) Two possible methods are based on $\mathrm{E}[y_i|\mathbf{x}_i] = \exp(\mathbf{x}_i'\boldsymbol{\beta})$ are
NLS of y_i on $\exp(\mathbf{x}_i'\boldsymbol{\beta})$ which minimizes $\sum_i (y_i - \exp(\mathbf{x}_i'\boldsymbol{\beta}))^2$.
MM estimation based on $\mathrm{E}[(y_i - \exp(\mathbf{x}_i'\boldsymbol{\beta}))\mathbf{x}_i] = \mathbf{0}$ which solves $\sum_i (y_i - \exp(\mathbf{x}_i'\boldsymbol{\beta}))\mathbf{x}_i = \mathbf{0}$.

(f) GMM estimator based on $\mathrm{E}[(y_i - \exp(\mathbf{x}_i'\boldsymbol{\beta}))\mathbf{z}_i] = \mathbf{0}$ which minimizes

$$Q_N(\boldsymbol{\beta}) = \frac{1}{N}\left(\sum_i (y_i - \exp(\mathbf{x}_i'\boldsymbol{\beta}))\mathbf{z}_i\right)' \mathbf{W}_N \left(\sum_i (y_i - \exp(\mathbf{x}_i'\boldsymbol{\beta}))\mathbf{z}_i\right),$$

where e.g. $\mathbf{W}_N = (\mathbf{Z}'\mathbf{Z})^{-1}$.

2.(a) Here need the probability

$$\Pr[y_i = 1|\mathbf{x}_i] = \Pr[y_i^* > 0] = \Pr[\mathbf{x}_i'\boldsymbol{\beta}+u_i > 0] = \Pr[-u_i > \mathbf{x}_i'\boldsymbol{\beta}] = \frac{1}{1+\exp(-\mathbf{x}_i'\boldsymbol{\beta})} = \frac{\exp(\mathbf{x}_i'\boldsymbol{\beta})}{1+\exp(\mathbf{x}_i'\boldsymbol{\beta})},$$

where use symmetry of density of u so that $-u$ has cdf $F(u) = 1/(1+\exp(-u))$.

(b) Letting $p_i = \Pr[y_i = 1|\mathbf{x}_i]$

$$\ln L(\boldsymbol{\beta}) = \sum\nolimits_i \ln f(y_i) = \sum\nolimits_i \ln[p_i^{y_i}(1-p_i)^{1-y_i}] = \sum\nolimits_i y_i \ln\left(\frac{\exp(\mathbf{x}_i'\boldsymbol{\beta})}{1+\exp(\mathbf{x}_i'\boldsymbol{\beta})}\right) + (1-y_i)\ln\left(\frac{1}{1+\exp(\mathbf{x}_i'\boldsymbol{\beta})}\right)$$

Obtain the MLE of $\boldsymbol{\beta}$ as solution to f.o.c. $\partial \ln L(\boldsymbol{\beta})/\partial\boldsymbol{\beta} = \mathbf{0}$.

(c) This is conditional logit. From memory [I did not ask for proof and only 1 point off if could not remember]

$$p_{ij} = \Pr[y_{ij} = 1] = \frac{\exp(\mathbf{x}_{ij}'\boldsymbol{\beta})}{\sum_{k=1}^{3}\exp(\mathbf{x}_{ik}'\boldsymbol{\beta})}, \quad j = 1, 2, 3,$$

and using the general form for the log-likelihood of multinomial models

$$\ln L = \sum\nolimits_i \ln f(y_i) = \sum\nolimits_i \ln\left(\prod\nolimits_{j=1}^{3} p_{ij}^{y_{ij}}\right) = \sum\nolimits_i \sum\nolimits_{j=1}^{3} y_{ij} \ln p_{ij} = \sum\nolimits_i \sum\nolimits_{j=1}^{3} y_{ij}\left(\ln \frac{\exp(\mathbf{x}_{ij}'\boldsymbol{\beta})}{\sum_{k=1}^{3}\exp(\mathbf{x}_{ik}'\boldsymbol{\beta})}\right)$$

(d) Independence of irrelevant alternatives assumption. Need to give some discussion as in class or notes mentioning red bus / blue bus problems or likely correlation of error ε_{ij} over j.

3.(a) In general

$$\mathrm{E}[y_2|\mathbf{x}, y_1^* > 0] = \mathrm{E}[\mathbf{x}_2'\boldsymbol{\beta}_2 + \varepsilon_2|\mathbf{x}, y_1^* > 0] = \mathbf{x}_2'\boldsymbol{\beta}_2 + \mathrm{E}[\varepsilon_2|\mathbf{x}, y_1^* > 0].$$

(b) Now

$$\mathrm{E}[y_2|\mathbf{x}, y_1^* > 0] = \mathbf{x}_2'\boldsymbol{\beta}_2 + \mathrm{E}[\delta\varepsilon_1 + \xi|\mathbf{x}, y_1^* > 0] = \mathbf{x}_2'\boldsymbol{\beta}_2 + \delta\mathrm{E}[\varepsilon_1|\mathbf{x}, y_1^* > 0].$$

as $\mathrm{E}[\xi|\mathbf{x}, y_1^* > 0] = \mathrm{E}[\xi] = 0$.

(c) Now, for notational simplicity suppressing dependence on $\mathbf{x}$

$$\mathrm{E}[y_2|y_1^* > 0] = \mathbf{x}_2'\boldsymbol{\beta}_2 + \delta\mathrm{E}[\varepsilon_1|\mathbf{x}_1'\boldsymbol{\beta}_1 + \varepsilon_1 > 0] = \mathbf{x}_2'\boldsymbol{\beta}_2 + \delta\mathrm{E}[\varepsilon_1|\varepsilon_1 > -\mathbf{x}_1'\boldsymbol{\beta}_1] = \mathbf{x}_2'\boldsymbol{\beta}_2 + \delta\frac{\phi(\mathbf{x}_1'\boldsymbol{\beta}_1)}{\Phi(\mathbf{x}_1'\boldsymbol{\beta}_1)}.$$

(d) Heckman two-step. First probit of y_1 on $\mathbf{x}_1$ – gives $\widehat{\boldsymbol{\beta}}_1$.
Then for $y_2 > 0$ do OLS of y_1 on $\mathbf{x}_2$ and $\phi(\mathbf{x}_1'\widehat{\boldsymbol{\beta}}_1)/\Phi(\mathbf{x}_1'\widehat{\boldsymbol{\beta}}_1)$ – gives $\widehat{\boldsymbol{\beta}}_2$ and $\widehat{\delta}$.

4. This question was answered extremely poorly by over half the class.

(a) Wald is usual z test in output.
Either use $p = 0.067 > 0.05$ so do not reject H_0
or use $z = 1.83 < 1.96$ so do not reject H_0.

(b) LR is given in the output.
Either use $p = 0.0666 > 0.05$ so do not reject H_0
or use $\mathrm{LR} = 3.36 < 1.96^2 = 3.84$ so do not reject H_0.

(c) There is not enough information to do the LM test.
Easiest to use NR_u^2 from auxiliary regression of y_i on $\partial \ln f(y_i)/\partial \boldsymbol{\theta}|_{\widehat{\boldsymbol{\theta}}_r}$ where $\boldsymbol{\theta}$ =$[\alpha, \beta]$ and $\widehat{\boldsymbol{\theta}}_r$ is the restricted MLE in the intercept only model.

5.(a) Here

$$\begin{aligned} & y_{it} = \alpha_i + \mathbf{x}_{it}'\boldsymbol{\beta}+\varepsilon_{it} \\ \Longrightarrow \quad & \bar{y}_i = \alpha_i + \bar{\mathbf{x}}_i'\boldsymbol{\beta} + \bar{\varepsilon}_i \text{ as } \bar{\alpha}_i = \alpha_i \\ \Longrightarrow \quad & y_{it} - \bar{y}_i = (\mathbf{x}_{it} - \bar{\mathbf{x}}_i)'\boldsymbol{\beta} + (\varepsilon_{it} - \bar{\varepsilon}_i) \end{aligned}$$

OLS of $y_{it} - \bar{y}_i$ on $(\mathbf{x}_{it} - \bar{\mathbf{x}}_i)$ gives

$$\widehat{\boldsymbol{\beta}}_{\text{FE}} = \left[\sum_{i=1}^{N} \sum_{t=1}^{T} (\mathbf{x}_{it} - \bar{\mathbf{x}}_i)(\mathbf{x}_{it} - \bar{\mathbf{x}}_i)' \right]^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} (\mathbf{x}_{it} - \bar{\mathbf{x}}_i)(y_{it} - \bar{y}_i).$$

(b) If fixed effects are present then FE estimator is consistent but RE is inconsistent. If random effects are present then both are consistent. Do Hausman test.
$H_0 : \alpha_i$ is random versus $H_a : \alpha_i$ is fixed. For $\boldsymbol{\beta}_1$ coefficients of time-varying regressors

$$\text{H} = \left(\widetilde{\boldsymbol{\beta}}_{1,\text{RE}} - \widehat{\boldsymbol{\beta}}_{1,\text{W}}\right)' \left[\widehat{\text{V}}[\widehat{\boldsymbol{\beta}}_{1,\text{W}}] - \widehat{\text{V}}[\widetilde{\boldsymbol{\beta}}_{1,\text{RE}}]\right]^{-1} \left(\widetilde{\boldsymbol{\beta}}_{1,\text{RE}} - \widehat{\boldsymbol{\beta}}_{1,\text{W}}\right),$$

and reject at level α if $\text{H} > \chi^2_\alpha(\dim[\boldsymbol{\beta}_1])$.

(c) In general H test uses $\widehat{\text{V}}[\widehat{\boldsymbol{\beta}}_{1,\text{W}} - \widetilde{\boldsymbol{\beta}}_{1,\text{RE}}]$. If $\widetilde{\boldsymbol{\beta}}_{1,\text{RE}}$ is fully efficient this simplifies to answer in (b). This is the case if $\varepsilon_{it} \sim \mathcal{N}[0,\sigma_\varepsilon^2]$. If this assumption cannot made need more general form for H. Implement by bootstrap.

(d) Given $y_{it} = \alpha_i + \mathbf{x}_{it}'\boldsymbol{\beta} + \varepsilon_{it}$, for the i^{th} individual stack all T observations, so

$$\begin{bmatrix} y_{i1} \\ \vdots \\ y_{iT} \end{bmatrix} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \alpha_i + \begin{bmatrix} \mathbf{x}_{i1}' \\ \vdots \\ \mathbf{x}_{iT}' \end{bmatrix} \boldsymbol{\beta} + \begin{bmatrix} \varepsilon_{i1} \\ \vdots \\ \varepsilon_{iT} \end{bmatrix}$$

or

$$\begin{aligned} \mathbf{y}_i &= \mathbf{e}\alpha_i + \mathbf{X}_i\boldsymbol{\beta} + \boldsymbol{\varepsilon}_i \\ \Longrightarrow \quad \mathbf{Q}\mathbf{y}_i &= \mathbf{Q}\mathbf{e}\boldsymbol{\alpha}_\mathbf{i} + \mathbf{Q}\mathbf{X}_i\boldsymbol{\beta} + \mathbf{Q}\boldsymbol{\varepsilon}_i \\ \Longrightarrow \quad \mathbf{Q}\mathbf{y}_i &= \mathbf{Q}\mathbf{X}_i\boldsymbol{\beta} + \mathbf{Q}\boldsymbol{\varepsilon}_i \end{aligned}$$

as $\mathbf{Qe} = \mathbf{I}_T - T^{-1}\mathbf{ee}']\mathbf{e} = \mathbf{e} - T^{-1}\mathbf{ee}'\mathbf{e} = \mathbf{e} - T^{-1}\mathbf{eT} = \mathbf{e} - \mathbf{e} = \mathbf{0}$.

The curve for this exam is only a guide. Course grade based on course score.

Scores out of	50		A	39 and above
$75th$ percentile	43	(84%)	A-	32 and above
Median	36	(72%)	B+	25 and above
$25th$ percentile	28	(56%)	B	18 and above